A Procedure for Refining a Coiled Coil Protein Structure Using X-Ray Fiber Diffraction and Modeling

Fatma Briki,* Jean Doucet,* and Catherine Etchebest†

*LURE, Bât 209D, Centre Universitaire Paris-Sud, F-91898 Orsay Cedex, France; †Bioinformatique Génomique et Moléculaire, INSERM U436, Université Paris 7, 75251 Paris Cedex 05, France

ABSTRACT We describe a combined use of experimental and simulation techniques to configure side chains in a coiled coil structure. As already demonstrated in a previous work, x-ray diffraction patterns from hard α-keratin fibers in the 5.15 Å meridian zone reflect the global configuration of the χ_1 dihedral angle of the coiled coil side chains. Molecular simulations, such as energy minimization and molecular dynamics, and rotameric representation in the PDB, are used here on a heterodimeric coiled coil to investigate the dihedral angle distribution along the sequence. Different procedures have been used to build the structure, the quality assessment was based on the agreement between the simulated diffraction patterns and the experimental ones in the fingerprint region of coiled coils (5.15 Å). The best one for building a realistic coiled coil structure consists of placing the side chains using molecular dynamics (MD) simulations, followed by side chain positioning using SMD or SCWRL procedures. The side chains and the backbone are equilibrated during the MD until they reach an equilibrium state for the t/g$^+$ ratio. Positioning the side chains on the resulting backbone, using the above procedures, gives rise to a well-defined 5.15 Å meridian reflection.

INTRODUCTION

Coiled coil structures are widespread motifs involved in a large number of oligomerization domains in various proteins. These structures are also the basic units at a molecular level in the fibrous proteins, such as intermediate filaments (Parry and Steinert, 1995). Regular coiled coil structures are obtained by two or more α-helices winding around each other (Crick, 1953b) to form a left-handed supercoil. Their stabilization is mainly achieved by interfacial hydrophobic interactions between residue side chains. The amino acid sequences composing these structures have a characteristic repetition of a seven-amino acid motif, usually noted $(\mathbf{abcdefg})_n$, with the **a** and **d** positions occupied by mostly hydrophobic residues (Parry et al., 1985; Parry and Fraser, 1985) facing inward to form a hydrophobic core into a knob-into-hole packing manner (Crick, 1953b). The side chains at the interface pack against each other in a specific configuration (O'Shea et al., 1991; Burkhard et al., 2000).

Until recently, the atomic structure of coiled coils was only known for short pieces involved in protein domains. The longest coiled coil structure determined at atomic resolution was that of the cFos-cJun bZip leucine zipper, which is 39 residues long (Glover and Harrison, 1995). In fibrous proteins, the coiled coil structures can be many heptads long. Our structural knowledge of this class of fibrous proteins is very limited because of the lack of adapted conditions for their crystallization to be studied by x-ray diffraction techniques; moreover, they are not soluble, so they cannot be studied by spectroscopic techniques such as NMR. High-resolution structural information on a long coiled coil piece has only been obtained recently by solving the crystalline structure of the 18 heptad repeat coiled coil domain of the actin bundling protein cortexillin (Burkhard et al., 2000).

Despite this lack in the fiber structures field, some information about these structures can be deduced from x-ray fiber diffraction. In the wide angle region, along the meridian axis (along the fiber axis), x-ray fiber diffraction patterns from hard α-keratins exhibit an intense reflection at 5.15 Å distance. This has been shown to arise essentially from the molecular structure. In a previous study we have demonstrated the side chain configuration dependence of the 5.15 Å reflection intensity (Busson et al., 1999). The origin of this reflection, which is the fingerprint on the diffraction pattern of the coiled coil structure in keratinous tissues, was unclear until this recent study. It was shown that an all-g$^+$ side chain configuration could switch off the intensity of the 5.15 Å reflection, whereas an all-*trans* configuration gives rise to a very intense reflection.

A fine analysis is, however, needed to complete this information, aiming at quantifying the proportion of the two main χ_1 (g$^+$ and *trans*) configurations in a coiled coil. Progress in homology modeling and protein design has generated considerable interest in methods for predicting side chains packing in hydrophobic cores of proteins (Bower et al., 1997; Dunbrack and Cohen, 1997; Tufféry et al., 1997; Koehl and Delarue, 1994). However, the present techniques are not practically validated because of the lack of an experimental tool able to provide information on this aspect. Moreover, most of the side chain configuration prediction studies are based on a statistical representation of the side chain conformations in the Protein Data Bank (PDB) (Bower et al., 1997; Tufféry et al., 1997). Actually, in their paper, Bower et al. developed a method for config-

Submitted August 14, 2001, and accepted for publication May 17, 2002.

Address reprint requests to Dr. Fatma Briki, LURE, Bât 209D, Centre Universitaire Paris-Sud, B.P. 34, F-91898 Orsay Cedex, France. Tel.: 33-1-64-46-88-20; Fax: 33-1-64-46-41-48; E-mail: briki@lure.u-psud.fr.

0006-3495/02/10/1774/10 $2.00

uring side chains on the basis of a rotameric library per residue (fewer than 10 per residue) extracted from the PDB and depending on the backbone parameters. Tufféry et al. used an approach in which a conformational search was performed in the rotameric space again built from the PDB structures. The effect of the backbone flexibility has also been analyzed using an approach consisting of small deviations of the structure from its experimental form; this study shows that only deviations larger than 2 Å RMSD give rise to significant changes on the predicted side chain configurations. However, all of these prediction methods are based on the rotameric and structural representation in the PDB. This criterion was seriously limiting for fibrous and membrane proteins, which are nearly absent in this database.

The objective of our study deals with the determination of a method for obtaining a realistic initial structure for molecular simulations on coiled coils. This is much needed for different studies in the field of fibrous proteins, like intermediate filaments, whose structures cannot be obtained from usual experimental techniques. Modeling the coiled coil structures is all the more important because these proteins are naturally submitted to different stresses whose effect is to induce structural changes that are hard to analyze experimentally. In addition, various mutations in the sequences affect the structure and the activity of these proteins (Kreis and Vale, 1999). We have already studied the effect of some stresses on the molecular structure of keratin (Kreplak et al., 2001), but an atomic structure is still lacking for understanding the behavior at this scale.

In the present work, we perform different simulations based on various techniques and force fields. Starting from an ideal backbone conformation, based on Crick equations (Crick, 1953a), the first simulation consisted of optimally placing the side chains with different side-chain positioning algorithms. In the second simulation, the side chains were placed in an all-g^+ χ_1 initial configuration, then the system was relaxed and explored through a short MD simulation monitored by the *trans* χ_1 configuration content. The third simulation consisted of using different positioning side chain algorithms on a backbone conformation selected along the dynamic trajectory. In all the cases the main purpose was to get the best agreement between the simulated diffraction patterns and the experimental one in the fingerprint region of the coiled coils (5.15 Å). In addition, the results were compared to the known high-resolution coiled coil crystallographic structures.

As a result, the side chain configurations can be inferred in the specific case of the coiled coil conformation, taking into account the dynamical aspect due to the backbone flexibility that had been shown to play an important role in side chain conformations (Harbury et al., 1995).

Two x-ray diffraction references have been used: 1) an experimental fiber diffraction pattern from hard α-keratin in hair; and 2) a calculated diffraction pattern from the crystallographic cortexillin structure, to avoid many extra signals originating from different parts of the hair fiber.

This study shows that the backbone flexibility, which plays a key role in the quality of the side chain positioning, can be achieved by a short molecular dynamics. For the side chain configuration we have used three positioning methods; two of them satisfy the x-ray criterion (SMD and SCWRL, see below). We show that the approach presented below can be followed for all coiled coils.

MATERIALS AND METHODS

Initial model building

A coiled coil dimer 101 residues long was built from the amino acid residues of part 1B of a type I 8C-1 (Dowling et al., 1986) and a type II 7C (Sparrow et al., 1989) keratin molecule. These have been shown experimentally to associate preferentially onto a heterodimer of the type I and the type II keratins (Herrling and Sparrow, 1991). The 1B piece of the molecule has been chosen because of its long uninterrupted heptad repeat sequence (18 heptads).

The coiled coil structure can be modeled by curves consisting of circles of radius r_2 processing with a frequency $\omega 1$ around a helix path of radius r_0 and a pitch p_0. The model parameterization was performed according to Busson and Doucet, 1999, using upgraded Crick's equations. These are based on a rotation matrix transformation relating a regular simple α-helix in the laboratory frame to a structure in a frame defined with its x axis on the radial vector and z axis tangent to a major helix. Therefore, a coiled helix was obtained by rotating a circle in the new frame when moving up the major helix path.

A value of $p_0 = 150$ Å has been attributed to the coiled coil pitch on the basis of experimental values for coiled coil pieces in the literature (O'Shea et al., 1991; Seo and Cohen, 1993). In this way, the trace of the C_α atoms was built and the other backbone atoms are placed according to a systematic construction with the O package (Jones, 1985). Side chains were then added with a chosen configuration. The dimeric molecule was built with two chains in exact register, as it has already been determined experimentally (Parry et al., 1985; Coulombe and Fuchs, 1990).

X-ray diffraction pattern calculation

X-ray diffraction patterns were simulated using a home-made program that calculates the scattered intensity on a plane detector by the model coiled coil. It determines successively for each pixel (x, z) of the simulated zone (x along the radial axis and z along the meridian) the corresponding scattering vector $\mathbf{S}$. The intensity $I(x, z)$ was then calculated as the square modulus of the structure factor $F(\mathbf{S})$ multiplied by the polarization and Lorentz factors. The polarization factor expression for a synchrotron radiation beam selected by a cylindrically curved crystal monochromator had been calculated (Kahn et al., 1982). The Lorentz factor, expressed as $(\cos 2\theta)^3$, where 2θ was the diffraction angle, accounts for both the beam divergence and the still geometry with a plane detector perpendicular to the x-ray beam.

The structure factor, which is a complex function, sums up the atomic contributions (phase and amplitude):

$$F(\mathbf{S}) = \sum_j \mathbf{f}_j\,(\mathbf{S}) e^{-2i\pi \mathbf{S}\cdot \mathbf{r}_j}$$

where $\mathbf{f}_j$ and $\mathbf{r}_j$ represent, respectively, the atomic scattering factor and the position vector for the jth atom. Simulations with 560 pixels along x and 140 along z have been performed assuming a wavelength $\lambda = 1.450$ Å and a sample-to-detector distance of 125 mm (experimental conditions for Fig. 1). The pixel size along the x and z axes was 0.1 mm, which corresponds

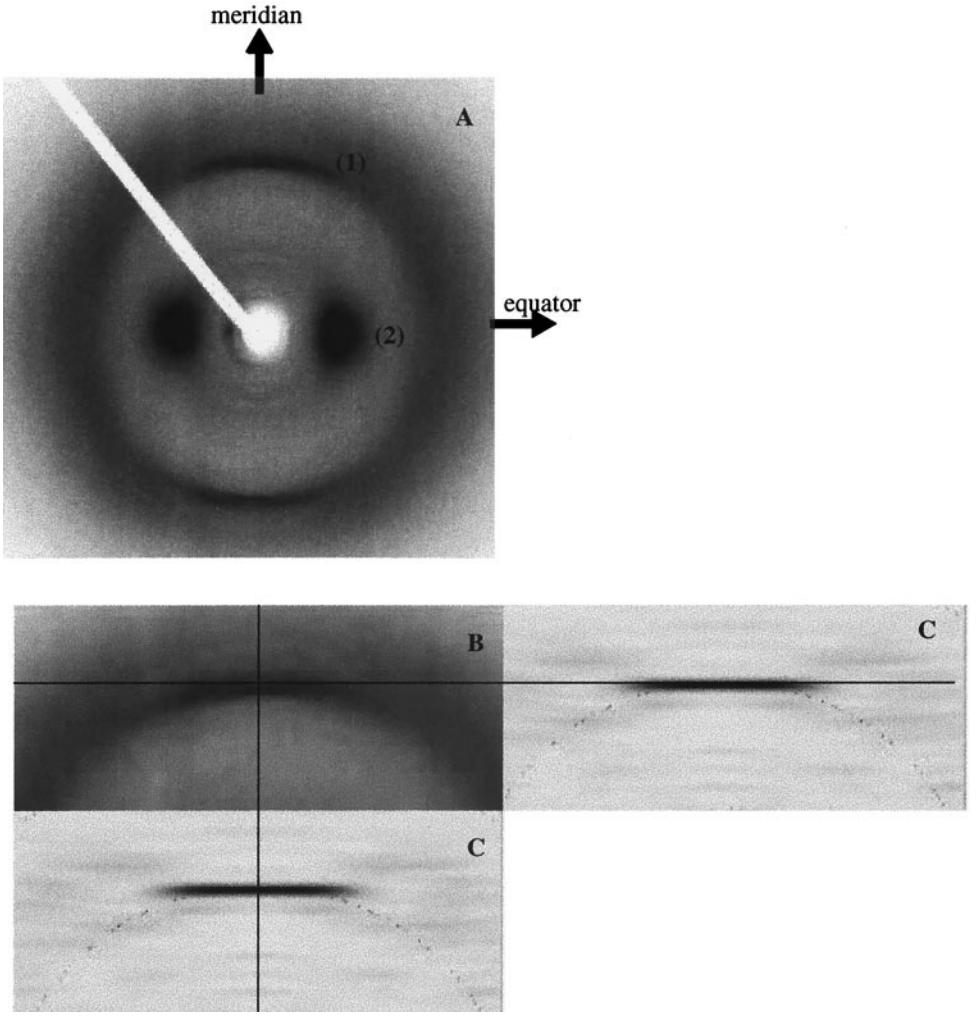

FIGURE 1 Experimental fiber diffraction pattern from hard α-keratin in hair sample (*A*) (the fiber axis is vertical, called meridian). (*1*) the arc at 5.15 Å spacing is directly related to the α-helix pitch projection along the coiled coil axis; (*2*) the broad spot at 9.8 Å on the normal to the axis direction (equator) corresponds to the mean distance between α-helical axes, for comparison, a blow-up of the 5.15 Å zone as used in the following calculated patterns is shown (*B*), along with a calculated pattern from an isolated experimental coiled coil (pdb code: 1D7M) (*C*); see the text for details. The dotted sharp arc on the simulated pattern indicates the 5 Å position.

in the reciprocal units to 5×10^{-4} Å^{-1}. Simulated patterns were finally cylindrically averaged by summing up the intensities produced by eight coiled coils related by successive 45° rotations around their axes to take into account the different orientations of coiled coils with respect to the x-ray beam. The graphic representation was simply achieved using gray scale images with a darkness proportional to the intensity.

Molecular simulations

Once the backbone model structure is built, the side chain dihedrals have to be chosen. The χ_1 dihedral angles are subjected to certain restrictions arising from steric hindrance between γ side chain atoms and the main chain atoms. Different procedures have been followed for the positioning and configuration of the side chains: use of the extended rotameric database from the PDB, energy minimization, and MD simulations.

Side chain positioning using a rotameric-based approach

Three approaches have been used to position the side chains on the coiled-coil backbone. All of them are based on a rotamer description for the side chain configurations, but they differ both by the rotamer database and the positioning algorithm. They are briefly summarized in the following paragraphs.

The first approach, called SMD (sparse matrix driven), developed by Tufféry et al., 1997, consists of a conformational search in the rotameric space using an energy criterion. The energy is computed using the complete FLEX force field (Lavery et al., 1986, 1995; Zakrzewska and Pullman, 1985). The rotameric space is constituted of 214 rotamers for the 20 amino acids. To overcome the combinatorial problem of side chain conformation prediction, the 3D structure is clustered into groups with a small number of strongly interacting residues. In each cluster, the set of energetically optimal rotamers is determined. The residue conformations are obtained through an appropriate combination of the different local solutions and the process is repeated two or three times, slightly modifying the definition of the clusters at each step. The final solution is the one corresponding to the minimal energy value. This procedure is systematically followed by a simplified energy minimization process in the dihedral conformational space of side chains. The residues are first ranked according to their interaction energy with the remaining residues, the first residue being that with the most unfavorable energy terms. Side chains are tested one at a time and the process is iterated until no interaction energy exceeds a given threshold (1 kcal/mol). All the side chains with energy conflicts are thus allowed to leave their rotamer state. At the end of the process the side chains are generally still close to their rotameric state, but do not present any energy clash.

The second approach (Koehl and Delarue, 1994), hereafter referred to as Confmat, uses the same backbone-independent rotamer library as Tufféry

et al. in the SMD method. The prediction of the side chain conformations is obtained through a mean field approximation. The mean field potential is defined as a sum of real potentials (only van der Waals interactions are considered) weighted by normalization factors. These factors are varied to yield a minimum of the effective potential of the effective system. The weighting factors are obtained from the elements of the conformational matrix. The procedure is first initialized with a conformational matrix CM_0, whose terms are the probability for each side chain to adopt a given rotamer. The mean potential is computed and the energies are then converted to probabilities defining a new conformational matrix CM_1 that is compared to CM_0. Then the process is iterated, modifying the CM_0 conformational matrix slightly at each step, taking into account the CM_1 values. When no significant difference exists between the CM_0 and CM_1 matrices, the convergence is considered to be reached and the process is stopped. The final solution corresponds, for each residue, to the rotamer with the highest probability. The side chains are then energy-minimized with the CHARMM force field in the Cartesian space with 1000 Powell steps.

The last approach (Bower et al., 1997) is called SCWRL. It uses a backbone-dependent rotamer library. The search strategy is made hierarchically. First, it chooses a structure with residues in their most probable backbone-dependent rotamers and systematically resolves the conflicts that arise from that structure, changing progressively to less probable rotamers until one is found without any conflict with the main chain. At this step, the side chain-side chains conflicts are examined. If residues clash, they are put into a cluster. The residues are allowed to explore all of their respective rotamers. When residues outside the cluster clash with the cluster residues for a given rotamer conformation, they are added to the cluster, leading to a progressive growth of the cluster size. At the end of the SCWRL procedure, the combination with the minimum steric clash score is assumed to be the best possible solution.

Configuring side chains using molecular dynamics

Because no simple criteria exist for attributing side chains in fibrous proteins, we have used the three most abundant χ_1 configurations (g^+, t, g^-) corresponding respectively to $-60°$, 180°, and 60° dihedral angles. A default choice was made to assign to the initial structure a g^+ χ_1 content of 82%, which is one of the default libraries in the O package. This rate results from the proportion of Val residues, which are all constructed in *trans* configuration, and Ser and Thr residues in g^- form. The other side chain dihedrals (χ_2, . . .) have been taken from the default library of the package. We note that the initial chosen structures are unfavorable, their optimization has to be achieved by energy minimization.

The structure was energy-minimized with the GROMOS (van Gunsteren and Berendsefn, 1987) package using a combination of 500 steepest descent steps followed by a 500-step conjugate gradient. Simulations were carried out in vacuo. The constant temperature MD calculations were carried out at $T = 300$ K, with an integration time step of 0.001 ps.

Two different protocols have been followed to achieve the dynamics. On the one hand, constraints were applied on the backbone (C_α, O, N, H) atoms during all the 0.4-ns molecular dynamics. This calculation (hereafter called "constrained-molecular dynamics") was achieved to save the initial global geometric parameters for the coiled coil (pitch, radius) and it also allowed the effect of these parameters on the conformational transitions of the side chains to be tested, as well as the influence of the geometry on these transitions. On the other hand, a molecular dynamics (hereafter called "free-molecular dynamics") has been carried out under different conditions: during the equilibrating period, the structure was harmonically constrained to its initial configuration with a force constant of 9000 kJ/mol·nm^2. The constraints were progressively reduced (1000 kJ/mol·nm^2/0.005 ns) until the temperature equilibrium was reached. When after 0.100 ns the equilibrium was reached, a free MD was then carried out for 0.4 ns. For the two protocols (free and restrained MD), two different MD calculations using different sets of initial velocities were accomplished.

Molecular dynamics and rotameric procedure combination

To take into account the backbone flexibility in the case of the use of a rotameric database, we used an approach combining an MD simulation for equilibrating the structure, followed by side chains positioning from a rotameric database). Structures from the free MD were picked up every 0.01 ns and the side chains placed using the SMD procedure, followed or not by energy minimization (see above).

RESULTS

In all the following simulations the structures have been shown to remain in a global coiled coil conformation whatever their statistical composition in side chain configuration. The global average geometric coiled coil parameters have only undergone small variations during the minimization steps. During the MD process the coiled coil pitch p_0 shifted progressively from 150 Å to an equilibrium average value of 140 Å.

Experimental information

Fig. 1 presents the experimental x-ray diffraction pattern obtained on the whole keratin fiber. The pattern shows a well-defined meridian reflection in the 5 Å zone, characteristic of both a coiled-coil conformation (Crick, 1953b) and a peculiar amino acid side chain configuration (Busson et al., 1999). We will focus all the analyses on the position, shape, and relative intensity of this reflection. It should be mentioned that the intense reinforcement of the 5.15 Å reflection is also associated with the supramolecular organization. This organization was not taken into account in the computed patterns on single molecules.

For ease of comparison, on the calculated diffraction patterns the intensity representation was selected to reveal the meridian 5.15 Å signal relative to the neighboring reflections, which are not seen on the experimental pattern because of the existence of a diffracted signal from other parts of the sample than the coiled coils. To help in reading and interpreting the calculated patterns, there are two relevant parameters: first, the meridian signal position near 5 Å (indicated by the central dashed arc), and second, the intensity of this reflection relative to the neighboring ones.

These facts are illustrated in Fig. 2 *A*, which was calculated on an all-g^+ structure that shows much more intense satellite reflections around the 5 Å position at the 5.15 Å fingerprint meridian. Clearly, this extinction is due to destructive interferences between the backbone atoms and the g^+ side chain atoms. Moreover, Fig. 2 *E*, which was calculated on an all-*trans* structure, shows constructive interferences between the backbone atoms and the *trans* side chains, which gives rise to the very intense signal at 5.15 Å. These facts have already been demonstrated in our previous paper (Busson et al., 1999).

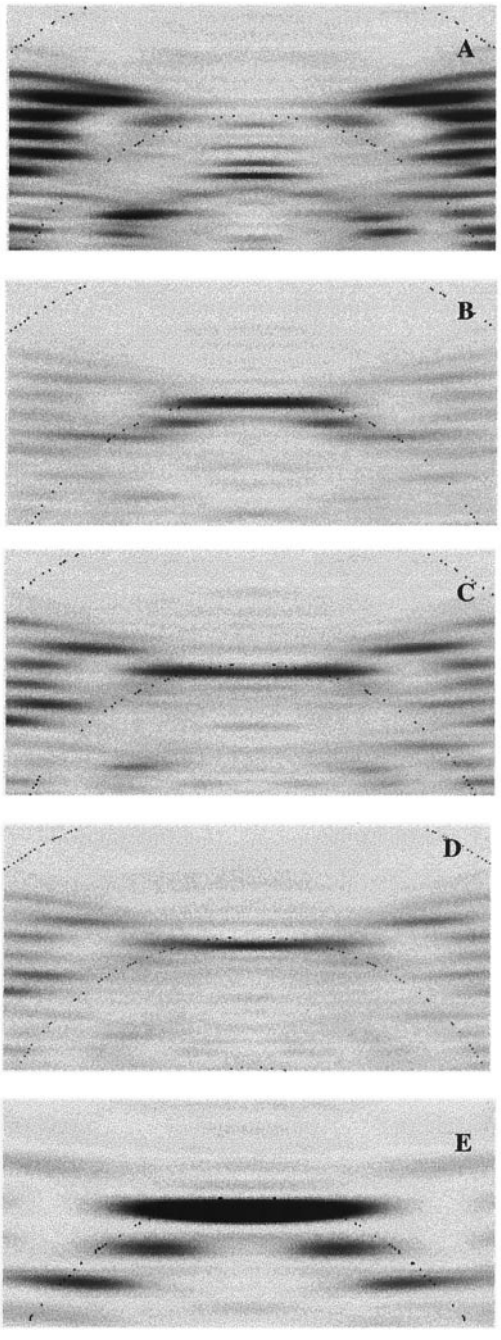

FIGURE 2 Calculated patterns built from structures using different side chain positioning methods on an ideal coiled coil backbone conformation, all-g^+ configuration of the side chains (*A*), side chains configured with SMD (*B*), Confmat (*C*), SCWRL (*D*), all-*trans* configuration (*E*). The dotted sharp arc indicates the 5 Å position.

Side chain positioning using rotameric databases and different algorithms on an ideal keratin backbone

Fig. 2 shows the diffraction patterns computed on the ideal backbone with the side chains positioned using the SMD procedure, the mean field procedure (Confmat), and the SCWRL procedure. For comparison, the pattern obtained from the structure with an all-g^+ side chain configuration is presented in Fig. 2 *A*. Clearly, no signal in the characteristic 5 Å zone is visible in this pattern. The structure obtained with the SMD procedure leads to a signal in the correct zone, but the ratio between the main reflection intensity compared to the secondary reflections is not satisfactory (Fig. 2 *B*); the intensity value is of the same order, instead of being significant at 5 Å. The structure with the mean-field approach leads to a different pattern from the previous one with a reflection split into two reflections in the 5 Å zone; this shape is rather different from the experimental one, which presents a unique well-centered maximum on the meridian axis. Moreover, the intensity is too low, not more significant than the neighboring satellite reflections (Fig. 2 *C*). The structure with the SCWRL procedure (Fig. 2 *D*) presents a shape similar to the one produced by the SMD method.

The statistical distribution of the χ_1 dihedral angle of the side chains is given for each pattern (Fig. 3). In the case of SMD, the proportion of *trans* configurations reaches ~50%. Analyzing the sequence, we notice that almost all the Glu, Lys, Arg, and Asp residues are in *trans* configurations regardless of their position in the heptad motif. In the case of the mean field approach, the proportion of *trans* configurations (~12%) was very small. For the SCWRL, the distribution was quite similar to that observed with SMD. The distribution of the side chains seems to be better for reproducing some important features of the diffraction pattern in the case of SMD and SCWRL compared to Confmat. Nevertheless, none of the structures produces a pattern sufficiently close to the experimental one. Energy-minimizing the global structure (backbone + side chains) after side chain positioning did not improve the pattern quality (data not shown).

The relevance of the rotameric databases for a coiled coil structure can be questioned. Indeed, the databases have been built up from the statistical data analysis from the PDB that does not include enough coiled-coil structure or fibrous proteins.

Side chain positioning with different procedures on the experimental backbone of the cortexillin coiled coil protein

To check the validity of the different approaches for a coiled coil structure, we selected the unique long high-resolution coiled coil structure available in the PDB, namely the cortexillin protein (code: 1D7M). Its simulated diffraction pattern is shown in Fig. 4 *A*. It displays well-defined features in the 5 Å zone, as for the keratin fiber. Starting from the experimental backbone, the side chains were positioned using the above procedures (Fig. 4, *B–D*). As in the previous case, the Confmat approach fails to give an appropriate diffraction signal, its shape was again different from the calculated pattern on the experimental structure, and its intensity was very low. The produced *trans* configuration

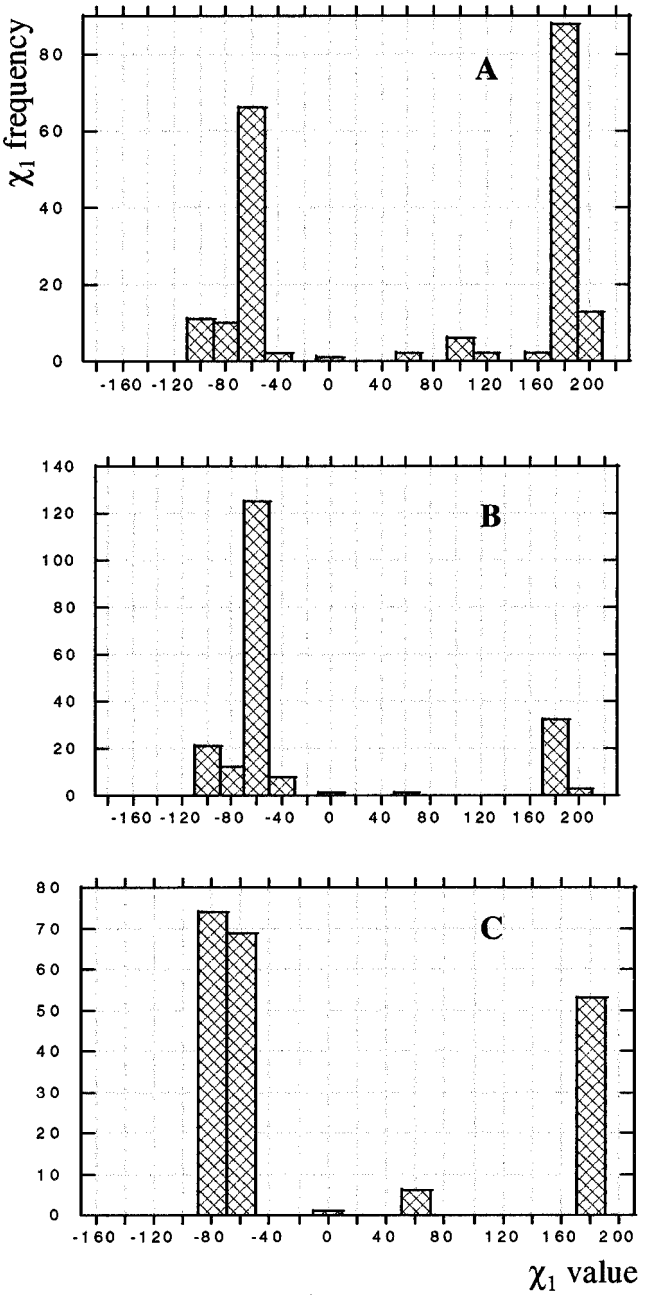

FIGURE 3 Calculated χ_1 distributions using different side chain positioning methods on an ideal coiled coil backbone conformation, SMD (*A*), Confmat (*B*), SCWRL (*C*).

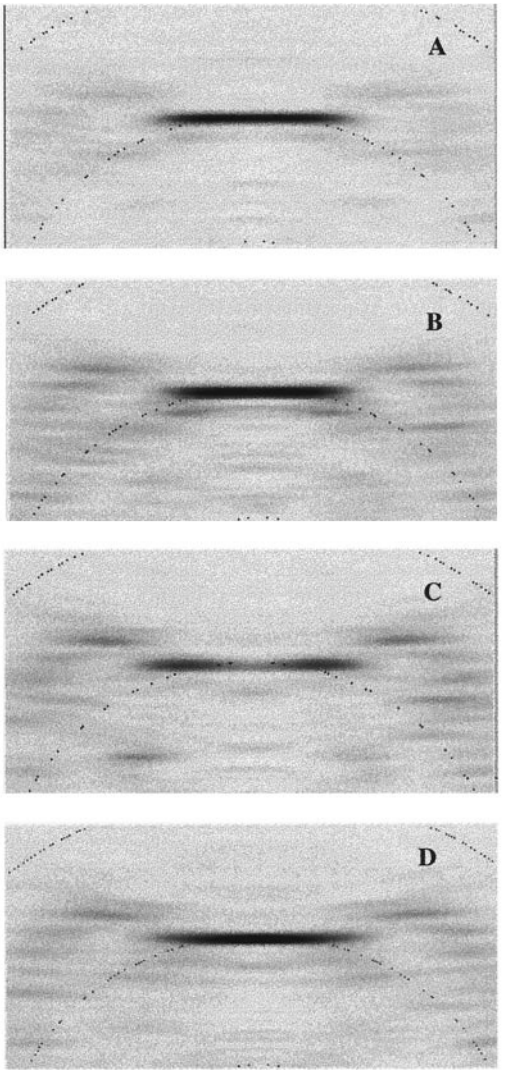

FIGURE 4 Calculated pattern from the experimental structure of the cortexillin coiled coil protein (pdb entry: 1D7M) (*A*), calculated patterns on the same backbone structure as in (*A*) but with the side chains configured using the different procedures (*B*) SMD, (*C*) Confmat, (*D*) SCWRL.

rate was again close to 16%, while it amounts to ~45% in the experimental structure. A slightly better signal was obtained with the SCWRL and SMD procedures: both the shape and the relative intensity of the 5 Å reflection were similar to the ones produced on the experimental crystallographic structure. Moreover, each of the three procedures produced better signal quality than they did on the ideal backbone keratin structure; the signal-to-noise ratio is improved as compared to Fig. 2 (see above).

Using this experimentally known structure, we have estimated the quality of the prediction of the method producing the best pattern. The quality of the prediction is defined as the number of side chains with χ_1 deviations $<30°$ from the experimental χ_1 value. According to this criterion, for all the residues in the structure the SMD procedure shows overall 54% correctly positioned side chains, with >77% side chains correctly positioned at the interface (**a**, **d** positions in the heptad repeat), and the SCWRL method predicts 57% residues in the right configuration with >81% at the interface. These results show that SMD and SCWRL produce very similar results, but SCWRL is slightly better.

Clearly, these approaches were able to place the side chains correctly at the interface of the backbone coiled coil conformation. For the other residues, it is likely that tran-

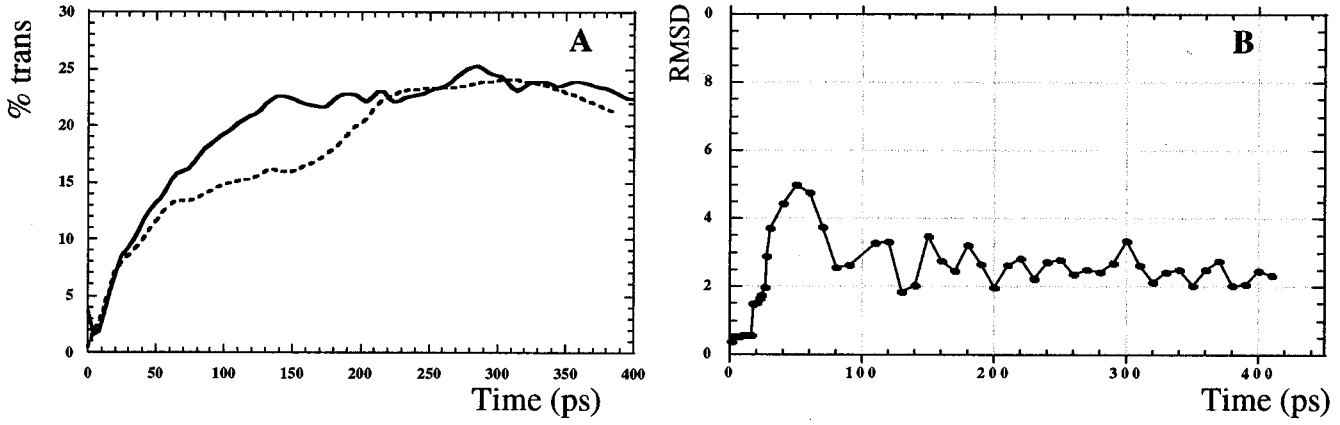

FIGURE 5 *trans* configuration χ_1 rate along two free molecular dynamics simulations from all-g^+ initial structure (*A*) and RMSD from the initial structure along the MD calculation (*B*).

sitions between different rotamers occur because of the lack in the representation of the environment.

The quality of the theoretical solution is assessed by comparison with the unique combination of rotamers provided by the x-ray experiment. Recent studies have mentioned the possible existence of a mixture of rotamers for a given structure. For instance, MacKenzie et al. (1996) have reported a fast exchange among χ_2 rotamers for all leucines in the glycophorin dimer studied with NMR experiments. Examination of the recently solved structure (MacKenzie et al., 1997) of the transmembrane part of the glycophorin dimer (code: 1AFO.pdb) indicates that first, the coiled coil extent is very short compared to the two fiber molecules examined here, and second, that most of the rotational flexibility of the leucine residues may be due to their occurrence in poorly constrained regions. In contrast, examination of the B factor values for the x-ray cortexillin dimer shows that most of the residues at the interface exhibit B values lower than the mean B value of the protein. This observation could indicate that in the cortexillin dimer, the interface residues are less flexible than the exposed ones. Thus, the probability of obtaining different combinations of rotameric states for the interface residues could be small.

Backbone structure influence

Aiming to test the backbone effect on the quality of the diffraction pattern, an MD simulation was carried out for a short period to allow backbone flexibility. All the side chains are initially in g^+ configuration. During the dynamics, the transitions were monitored using the increasing *trans* configuration rate. As we can see in Fig. 5 *A*, the free MD reaches an equilibrium state after ~0.2 ns, resulting in a rate of ~25% of the residues in *trans* configuration. In Fig. 5 the variations of the backbone RMSD compared to the initial ideal backbone B are indicated. When the equilibrium t/g^+ ratio was reached, one structure was selected at 0.21 ns. The dihedral distribution of the equilibrated system is shown in Fig. 6. Its backbone parameters are shown in Table 1. We observe for (φ, ψ) values an average shift of 5° from the ideal initial structure. The coiled coil helix pitch decreased from 150 Å to $p = 140$ Å. The decrease in the pitch value corresponds to an overwinding of the helices, but was still compatible with the observed pitches in coiled coil (Seo and Cohen, 1993).

From this MD structure a diffraction pattern has been computed (see Fig. 7 *A*). Compared to the previous patterns computed on the keratin molecule, we observe an improvement of the signal quality: the 5 Å arc is much more intense.

When the side chains are placed on this 0.21-ns MD backbone structure with the three side chain positioning procedures, all the calculated diffraction patterns are improved (see Fig. 7, *B–D*) compared to the patterns with the ideal backbone (Fig. 2). Compared to the pattern from MD without replacing the side chains, the quality is improved using side-chain placement algorithms.

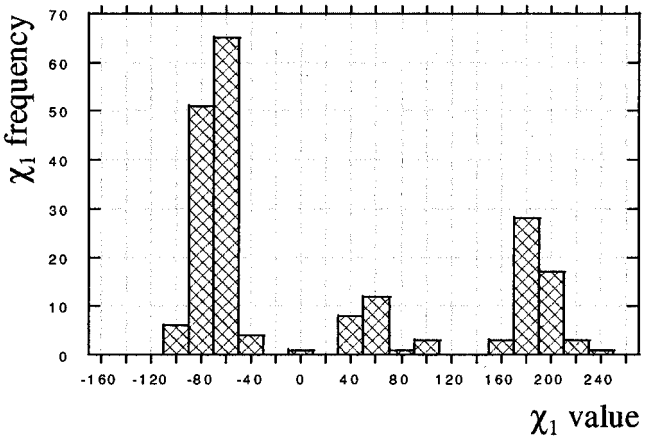

FIGURE 6 Coiled coil side chain χ_1 distribution from a selected molecular dynamics structure after the equilibration period (at 0.21 ns).

TABLE 1 Distribution of the backbone dihedral angles (degrees) in the calculated keratin structure, the GCN4 coiled coil, and the cortexillin fiber

	Mean φ	Std φ	Min	Max	Range	Mean ψ	Std ψ	Min	Max	Range
Keratin, initial backbone	−63.71	4.5	−75.4	−56.6	18.8	−41.7	4.9	−49.0	−27.6	21.4
0.21 ns-MD backbone	−58.68	10.0	−91.5	−24.7	66.8	−46.5	9.1	−70.0	−18.6	51.4
Cortexillin	−59.96	8.6	−83.4	−2.6	80.8	−44.8	9.3	−78.1	−16.5	61.6
GCN4	−62.20	7.4	−77.9	−37.5	40.4	−43.5	7.2	−63.7	−24.8	38.9

Note that the constrained backbone MD, which only allows small deviations to occur from the ideal geometry, gives rise to a simulated pattern presented in Fig. 8. The reflection at 5.15 Å is present, but is not well-contrasted relative to the other secondary reflections. Moreover, it presents three maxima instead of one, as is the case on the experimental pattern. All these results show that both the side chains and the backbone are important to provide a correct computed pattern.

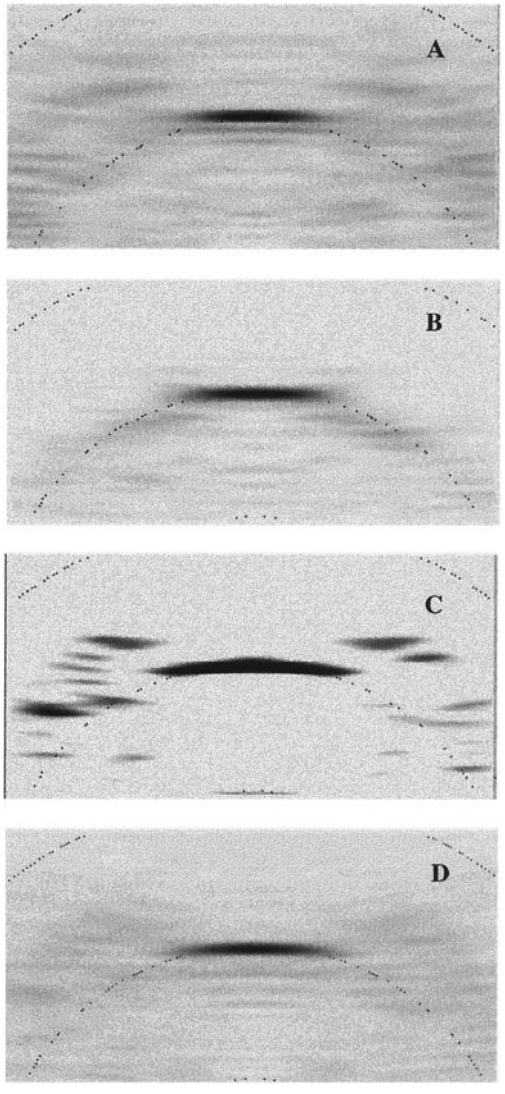

FIGURE 7 Calculated pattern from a selected MD structure 0.21 ns (*A*), calculated patterns on the same backbone structure as in (*A*) but with the side chains configured using the different procedures (*B*) SMD, (*C*) Confmat, (*D*) SCWRL.

DISCUSSION

Different methods for positioning the side chains on a given backbone have been proposed, but the main interest of our study was to use experimental information as a validating criterion. The main features that emerge from the present study concern two essential points: the side chain configuration and its coupling with an appropriate backbone representation. These two points are discussed below.

Accuracy of the positioning side chain procedure: rotamer database and algorithm effects

The three different procedures explored differ in two main criteria: the rotamer databases (SCWRL and SMD/Confmat) and the algorithm. The results obtained for the cortexillin have allowed us to assess their influence on the structure.

It seems that a fine-backbone-dependent database was not necessary in the present case due to the extreme regularity of the structure. In other words, the rotamers in the SCWRL database in the (φ, ψ) helical region are close to that defined in the average SMD database. On the contrary, the results obtained for the cortexillin show that the side chain positioning algorithm plays a key role. The SMD and SCWRL exhibit very similar χ_1 distribution (see Fig. 9) although based on a different positioning method. In the same way, Confmat and SMD, while using the same database, produce

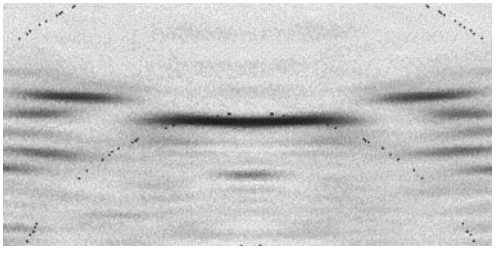

FIGURE 8 Calculated pattern from a selected structure after 0.21 ns backbone restrained molecular dynamics.

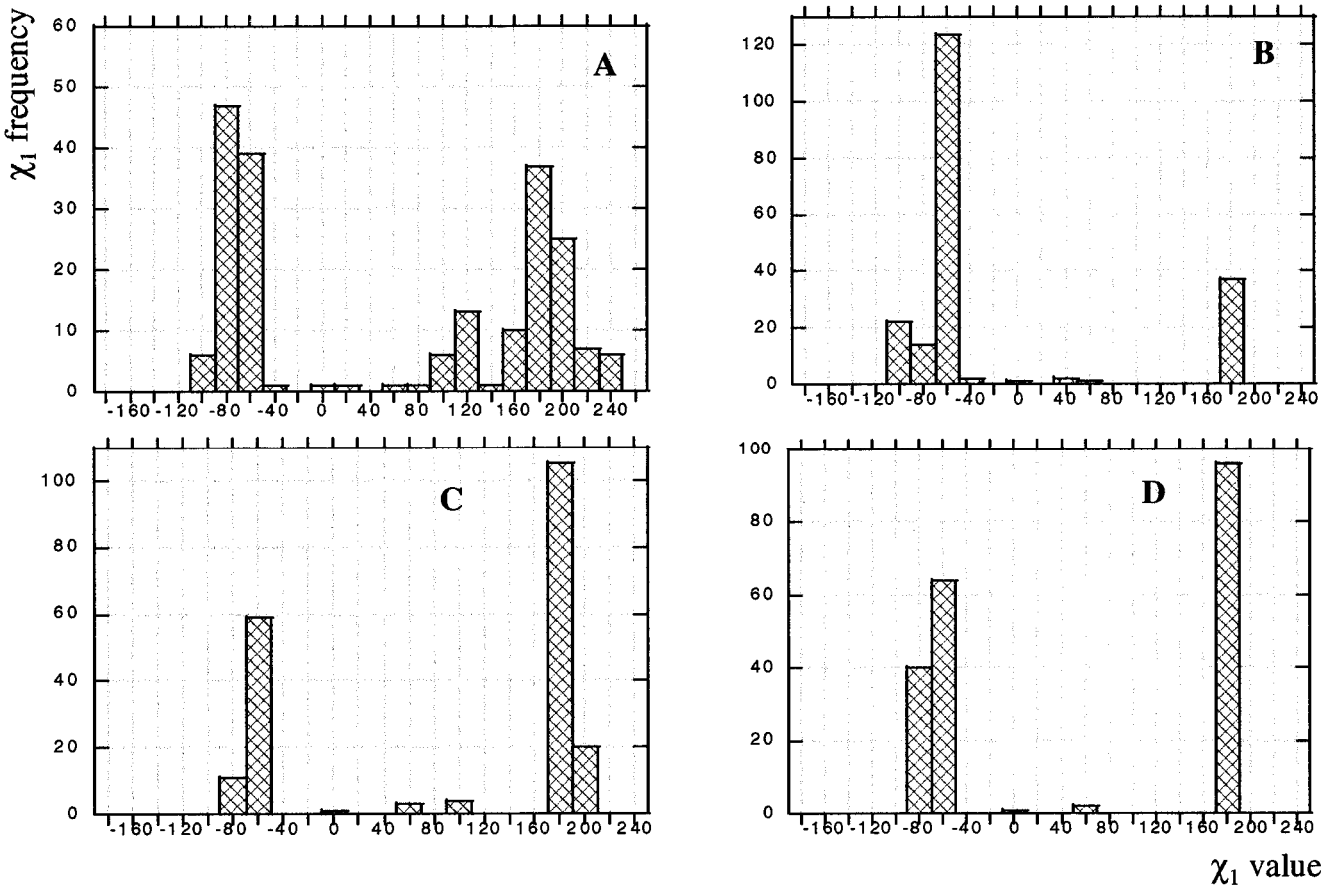

FIGURE 9 Coiled coil side chain χ_1 distribution from experimental cortexillin structure (*A*), side chains configured on the experimental cortexillin backbone using Confmat (*B*), SMD (*C*), SCWRL (*D*).

very different results. It may be noted that, although the different procedures use various simplified force fields, their roles do not seem crucial for determining the correct distribution of the side-chain angles.

Backbone influence

The different results obtained from the different starting structures show clearly that the backbone conformation is as important as the side chain conformations to provide a well-defined diffraction pattern. For instance, when the backbone conformation was restrained to its ideal conformation, letting only the side chains relax, the quality of the diffraction pattern was not improved.

At this point, it may be asked how a correct conformation for the backbone may be obtained. The simple procedure proposed, consisting of relaxing an initial "ideal" coiled coil structure with a short MD simulation, seems to be appropriate. The distribution of the (φ, ψ) dihedral angles at the end of the MD simulation is quite interesting. While the criterion used to select the final conformation concerns the g^+/t ratio of the side chain χ_1 angles, the distribution obtained for the backbone angles was quite similar to that observed in the crystallographic cortexillin dimer structure. In Table 1 the mean values, the standard deviations, the minimum and maximum along the sequence for the (φ, ψ) angles in keratin in the initial and final conformations, and the corresponding values for the cortexillin and the GCN4 coiled coils are shown. Clearly, the mean and the deviation for the keratin final structure are close to that of cortexillin. The fine analysis of the possible sequence dependence has not been performed because of the few data available. Nevertheless, it might be possible to avoid the MD step by selecting (φ, ψ) angles in the given distribution that appear characteristic of a long coiled coil structure. The GCN4 leucine zipper, which was considered as a template for a coiled coil structure, exhibits a slightly different distribution, closer to the ideal values chosen for constructing the initial backbone keratin structure.

This shows that the GCN4 backbone structure may not be appropriate for long fibrous proteins. Further studies need to be carried out to assess the influence of various parameters such as sequence specificity, sequence length, or environment on this coiled coil backbone structure.

CONCLUSIONS

In the present study we propose a new strategy combining a molecular simulation technique (molecular dynamics and side chain positioning procedure) with information from x-ray diffraction fiber patterns to obtain a detailed atomic coiled coil structure. The main result shows that a right side chain positioning method alone is not sufficient to provide correct diffraction patterns, an accurate backbone conformation is also necessary. The way to obtain both backbone and side chain conformations consists of selecting a short MD final structure and positioning the side chains with the SMD (or SCWRL) method using a rotamer database.

The present approach could be particularly interesting for most of the fiber proteins that are difficult to crystallize, but in contrast are easy to study with fiber diffraction techniques. Our strategy consists of selecting between putative coiled coil structures, built from various modeling methods, on the basis of an experimental criterion from x-ray fiber diffraction. Indeed, the present analysis shows that the keratin molecule is a good representative example of the ensemble of coiled coil structures. Its well-defined diffraction pattern could be used as a reference for other coiled coil proteins.

REFERENCES

Bower, M. J., F. E. Cohen, and R. L. Dunbrack, Jr. 1997. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* 267:1268–1282.

Burkhard, P., R. A. Kammerer, M. O. Steinmetz, G. P. Bourenkov, and U. Aebi. 2000. The coiled-coil trigger site of the rod domain of cortexillin I unveils a distinct network of interhelical and intrahelical salt bridges. *Structure Fold. Des.* 8:223–230.

Busson, B., F. Briki, and J. Doucet. 1999. Side-chain configurations in coiled coils revealed by the 5.15-A meridional reflection on hard alpha-keratin x-ray diffraction patterns. *J. Struct. Biol.* 125:1–10.

Busson, B., and J. Doucet. 1999. Modeling alpha-helical coiled coils: analytic relations between parameters. *J. Struct. Biol.* 127:16–21.

Coulombe, P. A., and E. Fuchs. 1990. Elucidating the early stages of keratin filament assembly. *J. Cell Biol.* 111:153–69.

Crick, F. H. C. 1953a. The Fourier transform of a coiled coil. *Acta Crystallogr.* 6:685–689.

Crick, F. H. C. 1953b. The packing of alpha-helices: simple coiled coils. *Acta Crystallogr.* 6:689–697.

Dowling, L. M., W. G. Crewther, and A. S. Inglis. 1986. The primary structure of component 8c-1, a subunit protein of intermediate filaments in wool keratin. Relationships with proteins from other intermediate filaments. *Biochem. J.* 236:695–703.

Dunbrack, R. L., Jr., and F. E. Cohen. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* 6:1661–1681.

Glover, J. N., and S. C. Harrison. 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature.* 373:257–261.

Harbury, P. H., B. Tidor, and P. Kim. 1995. Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proc. Natl. Acad. Sci. U.S.A.* 92:8408–8412.

Herrling, J., and L. G. Sparrow. 1991. Interactions of intermediate filament proteins from wool. *Intern. J. Biol. Macromol.* 13:115–119.

Jones, T. A. 1985. Interactive computer graphics. *Methods Comp. Methods Enzymol.* 115:157–171.

Kahn, R., R. Fourme, A. Gader, J. Janin, C. Dumas, and D. Andre. 1982. Macromolecular crystallography with synchrotron radiation: photographic data collection and polarization correction. *J. Appl. Crystallogr.* 15:330–337.

Koehl, P., and M. Delarue. 1994. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* 239:249–275.

Kreis, T., and R. Vale. 1999. Guidebook to the Cytoskeletal and Motor Proteins. Oxford University Press, Oxford.

Kreplak, L., J. Doucet, and F. Briki. 2001. Unraveling double stranded alpha-helical coiled coils: an x-ray diffraction study on hard alpha-keratin fibers. *Biopolymers.* 58:526–533.

Lavery, R., I. Parker, and J. Kendrick. 1986. A general approach to the optimization of the conformation of ring molecules with an application to valinomycin. *J. Biomol. Struct. Dyn.* 4:443–462.

Lavery, R., K. Zakrzewska, and H. Sklenar. 1995. Junction minimization of nucleic acids. *Comp. Phys. Commun.* 91:135–138.

MacKenzie, K. R., J. H. Prestegard, and D. M. Engelman. 1996. Leucine side-chain rotamers in a glycophorin A transmembrane peptide as revealed by three-bond carbon-carbon couplings and 13C chemical shifts. *J. Biomol. NMR.* 7:256–260.

MacKenzie, K. R., J. H. Prestegard, and D. M. Engelman. 1997. A transmembrane helix dimer: structure and implications. *Science.* 276: 131–133.

O'Shea, E. K., J. D. Klemm, P. S. Kim, and T. Alber. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science.* 254:539–544.

Parry, D. A. D., and R. D. B. Fraser. 1985. Intermediate filament structure. 1. Analysis of IF protein sequence data. *Intern. J. Biol. Macromol.* 7:203–213.

Parry, D. A. D., and P. M. Steinert. 1995. Intermediate Fiilament Structure. Springer Verlag, Heidelberg, Germany.

Parry, D. A. D., A. C. Steven, and P. M. Steinert. 1985. The coiled coil molecules of intermediate filaments consist of two parallel chains in exact axial register. *Biochem. Biophys. Res. Commun.* 127:1012–1018.

Seo, J., and C. Cohen. 1993. Pitch diversity in alpha-helical coiled coils. *Proteins: Struct., Funct., Genet.* 15:223–234.

Sparrow, L. G., C. P. Robinson, D. T. W. MacMahon, and M. R. Rubira. 1989. The amino acid sequence of component 7c, a type II intermediate filament protein from wool. *Biochem. J.* 261:1015–1022.

Tufféry, P., C. Etchebest, and S. Hazout. 1997. Prediction of protein side chain conformations: a study on the influence of backbone accuracy on conformation stability in the rotamer space. *Protein Eng.* 10:361–372.

van Gunsteren, W. F., and H. J. C. Berendsefn. 1987. Groningen Molecular Simulation (GROMOS), the Netherlands.

Zakrzewska, K., and A. Pullman. 1985. Optimized monopole expansions for the representation of the electrostatic properties of polypeptide and proteins. *J. Comp. Chem.* 6:265–273.